

VARIABLE ELIMINATION IN POST-TRANSLATIONAL MODIFICATION REACTION NETWORKS WITH MASS-ACTION KINETICS

ABSTRACT. We define a subclass of Chemical Reaction Networks called Post-Translational Modification systems. Important biological examples of such systems include MAPK cascades and two-component systems which are well-studied experimentally as well as theoretically. The steady states of such a system are solutions to a system of polynomial equations with as many variables as equations. Even for small systems the task of finding the solutions is daunting. We develop a mathematical framework based on the notion of a *cut*, which provides a linear elimination procedure to reduce the number of variables in the system. The steady states are parameterized algebraically by a set of “core” variables, and the non-negative steady states correspond to non-negative values of the core variables. Further, minimal cuts are the connected components in the species graph and provide conservation laws. A criterion for when a set of independent conservation laws can be derived from cuts is given.

Keywords: Polynomial equations, Mass-action kinetics, MAPK cascade, Rational functions, Chemical Reaction Networks

1. INTRODUCTION

Signaling systems play an important role in regulation of cellular processes and are essential for cellular decision making. Typical signaling systems react to stimulus in the (cellular) environment and transmit a signal through connected layers of biochemical species. The layers provide means to adjust the response according to the stimulus. A common form of signaling systems is Post-translational Modification (PTM) systems where species are activated in chemical reactions in order to propagate the signal through the system.

PTM systems have attracted considerable theoretical attention due to their abundance in nature [12] and regular form [17]. The dynamics can be modeled as $\frac{dx(t)}{dt} = p(x)$, where $x = (x_1, \dots, x_n)$ are the variables (concentrations of species) of the system and $p(x)$ is a vector of polynomials in x . Only certain types of reactions are allowed, restricting the form of $p(x)$. In particular, small specific systems have been scrutinized, focusing on the dynamical behavior and the steady states of the systems. Examples include the biologically important MAPK cascades [12, 13, 15], as well as simpler signaling cascades [9, 11, 20].

We focus on the steady states of a PTM system (defined formally in the next section) and how to determine them. Taken with mass-action kinetics, the system’s steady states are solutions to a set of polynomial equations in the species and with coefficients given by unknown kinetic rates (i.e. unspecified parameters). In particular, the number of equations

Date: 19/07/2011.

Authors affiliation: Bioinformatics Research Centre, Aarhus University, C. F. Møllers Allé 8, DK-8000 Aarhus, Denmark.

Corresponding author: Elisenda Feliu, efeliu@birc.au.dk.

to be solved is equal to the number of species. Even small systems might have many variables such that analytical solutions are difficult to obtain and numerical solutions are prone to errors. Further, many PTM systems admit multistationarity (the existence of more than one steady state under particular biological conditions) which is a mechanism for cellular decision making [18]. It is therefore of interest to determine the parameters for which mono- and multistationarity occur. Several non-necessary conditions for a unique positive steady state are known [1, 4, 7], but when these fail, multistationarity is difficult to determine and often decided based on a random parameter search. Procedures to eliminate variables (hence, equations) is therefore fundamental to the theoretical understanding of these systems as well as for numerical analysis.

Our work is inspired by previous work by Thomson and Gunawardena (TG) [17] which we extend to embrace a range of important PTM systems such as signaling cascades (including the MAPK cascade) and two-component systems with phosphorelays and phosphotransfer [14], as well as systems with self-interactions. We develop the idea of a *cut* \mathcal{S}_α , a subset of the substrates \mathcal{S} with certain properties that allow us to express the steady state equations as rational functions in the “core” variables $\mathcal{S} \setminus \mathcal{S}_\alpha$, providing an algebraic parameterization of the steady states in terms of the core variables. If the core variables take positive values at steady state, then we show that all other concentrations are either zero or positive as well.

Further, we show that cuts relate to *conservation laws* (conserved quantities that imply that the dynamics takes place in an affine invariant subspace of \mathbb{R}^n) that arise as connected components in the species graph [1]. Conservation laws are often used as a first step to reduce the dimensionality of the system. In our approach, conservation laws come into play after elimination of variables from the steady state equations. In this way, we allow for a larger reduction in the number of core variables.

Our approach makes use of algebraic tools as well as some basic graph properties; for example Tutte’s Matrix-Tree theorem [19, 17]. One benefit is that parameters are treated as symbolic constants and do not need to be fixed or assumed known in advance. This is particularly relevant in biology, where we often are faced with systems that depend on experimental parameters (kinetic rates), which are difficult to determine.

2. POST-TRANSLATIONAL MODIFICATION SYSTEMS

2.1. PTM system. A *post-translation modification (PTM) system* consists of two non-empty sets of species, $\mathcal{S} = \{S_1, \dots, S_N\}$ (the *substrates*) and $\mathcal{Y} = \{Y_1, \dots, Y_P\}$ (the *intermediate complexes*) with $\mathcal{S} \cap \mathcal{Y} = \emptyset$, and a set of reactions $\text{Rct} = R_a \cup R_b \cup R_c \cup R_d$ with associated positive reaction rate constants:

$$\begin{aligned} R_a &= \{S_i + S_j \xrightarrow{a_{i,j}^k} Y_k \mid (i, j, k) \in I_a\} & R_c &= \{Y_i \xrightarrow{c_{i,j}} Y_j \mid (i, j) \in I_c, i \neq j\} \\ R_b &= \{Y_k \xrightarrow{b_{i,j}^k} S_i + S_j \mid (i, j, k) \in I_b\} & R_d &= \{S_i \xrightarrow{d_{i,j}} S_j \mid (i, j) \in I_d, i \neq j\} \end{aligned}$$

for $I_a, I_b \subseteq \{1, \dots, N\}^2 \times \{1, \dots, P\}$, $I_c \subseteq \{1, \dots, P\}^2$ and $I_d \subseteq \{1, \dots, N\}^2$. To fix the notation, we assume that any $(i, j, k) \in I_a \cup I_b$ satisfies $i \leq j$, so that self-interactions a priori are allowed. If the rate constants are not required, we put an arrow to indicate a reaction and omit the rates. Further:

- (i) All chemical species are involved in at least one reaction.
- (ii) For every intermediate complex Y_k there exist $i \leq j$, indices k_1, \dots, k_r and a chain of reactions $Y_k \rightarrow Y_{k_1} \rightarrow \dots \rightarrow Y_{k_r} \rightarrow S_i + S_j$.

Assumption (ii) ensures that Y_k *ultimately dissociates into two substrates*. Also, we allow that there are more than one Y_k such that $S_i + S_j \rightarrow Y_k$ or $Y_k \rightarrow S_i + S_j$ for given S_i, S_j . For convenience, we put $c_{i,j} = 0$, $d_{i,j} = 0$ if $(i, j) \notin I_c$ or I_d respectively, and similarly $a_{i,j}^k = 0$ and $b_{i,j}^k = 0$ if $(i, j, k) \notin I_a$ or I_b , respectively. For $i \leq j$ and k , we define $a_{j,i}^k = a_{i,j}^k$ and $b_{j,i}^k = b_{i,j}^k$. For later use, we define

$$\mathcal{S}_\circ = \{S_i \in \mathcal{S} \mid (i, i, k) \in I_a \cup I_b \text{ for some } k\}$$

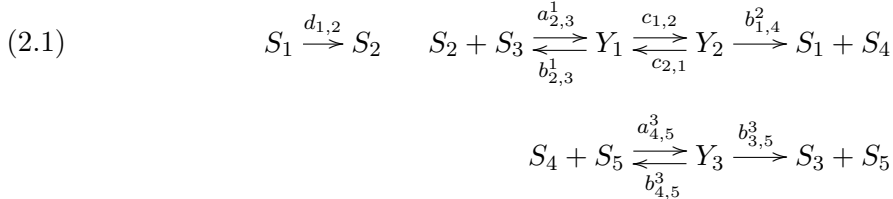
to be the set of self-interacting substrates.

This setting fits post-translational modification of proteins catalyzed by enzymes as well as the transfer of modifier groups:



where S^*, P^* are modified proteins (substrates), S, P their corresponding unmodified forms, E an enzyme (substrate) and Y an intermediate complex. That is, $\mathcal{S} = \{S, S^*, P, P^*, E\}$ and $\mathcal{Y} = \{Y\}$. In the first case, the attachment of the modifier group is catalyzed by the enzyme E , whereas in the second case, a modifier group is transferred from P^* to S . Modification of a substrate or an intermediate complex without the involvement of other species is modeled by $S \rightarrow S^*$ and $Y \rightarrow Y^*$, respectively.

As an example consider the PTM system with $\mathcal{S} = \{S_1, S_2, S_3, S_4, S_5\}$, $\mathcal{Y} = \{Y_1, Y_2, Y_3\}$ and reactions



One interpretation is that S_1 is modified to S_2 . The modifier group is then transferred from S_2 to S_3 with the formation of two intermediate complexes Y_1, Y_2 , causing the modification of S_3 to S_4 and the demodification of S_2 to S_1 . Finally, S_4 is demodified via a Michaelis-Menten mechanism catalyzed by an enzyme S_5 .

Nomenclature. We introduce a few concepts that will be used in the following, some of which are taken from Chemical Reaction Network Theory (CRNT) [6, 8]. Consider the set of *complexes* of the reaction system:

$$\mathcal{C} = \mathcal{Y} \cup \{S_i, S_j \mid (i, j) \in I_d\} \cup \{S_i + S_j \mid (i, j, k) \in I_a \cup I_b \text{ for some } k\}.$$

Then:

- $A \in \mathcal{C}$ *reacts to* $B \in \mathcal{C}$ if there exists a reaction $A \rightarrow B$.
- $A \in \mathcal{C}$ *ultimately reacts to* $B \in \mathcal{C}$ if there exists a sequence of reactions $A \rightarrow A_1 \rightarrow \dots \rightarrow A_r \rightarrow B$ with $A_m \in \mathcal{C}$. If $A_m \in \tilde{\mathcal{Y}} \subseteq \mathcal{Y}$ for all m , then A *ultimately reacts to* B *via* $\tilde{\mathcal{Y}}$.
- S_i and S_j *interact* if for some Y_k either $S_i + S_j$ reacts to Y_k or vice versa.

- S_i, S_j are 1-linked if $d_{i,j}$ or $d_{j,i} \neq 0$. Y_k, Y_v are 1-linked if $c_{k,v}$ or $c_{v,k} \neq 0$. S_i and Y_k are 1-linked if for some j , $S_i + S_j$ reacts to Y_k or vice versa ($j = i$ is allowed).

Assumption (ii) of a PTM system ensures that all intermediate complexes ultimately react to some $S_i + S_j$ via \mathcal{Y} .

2.2. Mass-action kinetics. The set of reactions together with their associated rate constants give rise to a polynomial system of ordinary differential equations taken with *mass-action kinetics*:

$$\begin{aligned} \dot{Y}_k &= \sum_{j=1}^N \sum_{i=1}^j (a_{i,j}^k S_i S_j - b_{i,j}^k Y_k) + \sum_{v=1}^P (c_{v,k} Y_v - c_{k,v} Y_k), & k = 1, \dots, P, \\ \dot{S}_i &= \sum_{j=1}^N \sum_{k=1}^P \epsilon_{i,j} (-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j=1}^N (d_{j,i} S_j - d_{i,j} S_i), & i = 1, \dots, N, \end{aligned}$$

where $\epsilon_{i,j} = 1$ if $i \neq j$ and 2 if $i = j$ and where \dot{x} denotes dx/dt for $x = x(t)$. Here we abuse notation and let S_i, Y_k denote the concentrations of the species S_i, Y_k as well. The steady states are the solutions to the polynomial system obtained by setting the derivatives to zero, i.e. $\dot{Y}_k = 0$ and $\dot{S}_i = 0$:

$$(2.2) \quad 0 = \sum_{j=1}^N \sum_{i=1}^j (a_{i,j}^k S_i S_j - b_{i,j}^k Y_k) + \sum_{v=1}^P (c_{v,k} Y_v - c_{k,v} Y_k), \quad k = 1, \dots, P,$$

$$(2.3) \quad 0 = \sum_{j=1}^N \sum_{k=1}^P \epsilon_{i,j} (-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j=1}^N (d_{j,i} S_j - d_{i,j} S_i), \quad i = 1, \dots, N.$$

This system is quadratic in the variables Y_k, S_i , but the only quadratic terms are of the form $S_i S_j$. It is linear in Y_k .

It is convenient to treat the reaction rate constants as parameters with unspecified (positive) values and view $a_{i,j}^k, b_{i,j}^k, c_{u,v}, d_{w,t}$ as symbols. For that, let

$$\text{Con} = \{a_{i,j}^k | (i, j, k) \in I_a\} \cup \{b_{i,j}^k | (i, j, k) \in I_b\} \cup \{c_{k,v} | (k, v) \in I_c\} \cup \{d_{k,v} | (k, v) \in I_d\}$$

be the set of the non-zero parameters (symbols). Then, the system (2.2)-(2.3) is quadratic in $\mathcal{S} \cup \mathcal{Y}$ with coefficients in the field $\mathbb{R}(\text{Con})$. Further, if all S_i are considered part of the coefficient field, then the system is linear with coefficients in $\mathbb{R}(\text{Con} \cup \mathcal{S})$ and variables Y_1, \dots, Y_P .

Only *non-negative solutions* of the steady state equations are biologically meaningful. To study positivity of solutions, we introduce the concept of *S-positivity*. Let $X = \{x_1, \dots, x_r\}$ be a finite set. A non-zero polynomial in $\mathbb{R}[X]$ with non-negative coefficients is called *S-positive*. Similarly, a rational function f is S-positive if it is a quotient of two S-positive polynomials. If x_1, \dots, x_r are substituted by positive real numbers in f , we obtain a positive real number. In general, a rational function $f = p/q$ in z_1, \dots, z_s and coefficients in $\mathbb{R}(X)$ is S-positive if the coefficients of p and q are S-positive rational functions in x_1, \dots, x_r . If f is a rational function in x_1, \dots, x_r and $x_i = g(x_1, \dots, \hat{x}_i, \dots, x_r)$ with g a rational function, then substituting g into f gives f as a rational function in $x_1, \dots, \hat{x}_i, \dots, x_r$.

The differential equations of Example (2.1) are:

$$\begin{aligned}
(2.4) \quad \dot{Y}_1 &= a_{2,3}^1 S_2 S_3 - (b_{2,3}^1 + c_{1,2}) Y_1 + c_{2,1} Y_2 & \dot{S}_1 &= -d_{1,2} S_1 + b_{1,4}^2 Y_2 \\
\dot{Y}_2 &= c_{1,2} Y_1 - (b_{1,4}^2 + c_{2,1}) Y_2 & \dot{S}_2 &= d_{1,2} S_1 - a_{2,3}^1 S_2 S_3 + b_{2,3}^1 Y_1 \\
\dot{Y}_3 &= a_{4,5}^3 S_4 S_5 - (b_{4,5}^3 + b_{3,5}^3) Y_3 & \dot{S}_3 &= -a_{2,3}^1 S_2 S_3 + b_{2,3}^1 Y_1 + b_{3,5}^3 Y_3 \\
\dot{S}_5 &= -a_{4,5}^3 S_4 S_5 + (b_{4,5}^3 + b_{3,5}^3) Y_3 & \dot{S}_4 &= -a_{4,5}^3 S_4 S_5 + b_{1,4}^2 Y_2 + b_{4,5}^3 Y_3.
\end{aligned}$$

To compute the steady states, we can use $\dot{Y}_3 = 0$ to eliminate Y_3 as a function of the substrates. Also Y_1, Y_2 can be eliminated by solving the linear system $\dot{Y}_1 = \dot{Y}_2 = 0$. This is a general feature of PTM systems and is covered in Section 3.1.

Further, observe that $\dot{S}_5 + \dot{Y}_3 = 0$, which implies that the sum $S_5 + Y_3$ is independent of time and thus conserved. In fact, it implies that one of the equations $\dot{Y}_3 = 0$ and $\dot{S}_5 = 0$ is redundant. Removing one of them leaves a polynomial system with 7 equations in 8 variables, and thus the solutions to the steady state equations form an algebraic variety of dimension at least one. This redundancy can be compensated for by fixing the value $S_5 + Y_3 = \bar{S}$ and adding this relation to the steady state equations.

In the next section we discuss the existence of the so-called *conservation laws* and provide a graphical procedure to determine (some of) them. In most cases the procedure provides a set of independent conservation laws, but, as will be discussed below, this might not always be the case.

2.3. Conservation laws. We consider systems where inflow of species is not allowed and species are not degraded or able to diffuse out. Such systems are “entrapped” in contrast to open systems (so-called “continuous flow stirred tank reactors”) [3]. PTM systems are entrapped and have conservation laws that reflect that the total amount of species remains constant either in free form S_i or in bounded form Y_j . These laws follow from the system of differential equations and appear as linear combinations of species (e.g. $S_5 + Y_3 = \bar{S}$ in the example above).

The existence of conservation laws implies that the dynamics of the system takes place in a proper invariant subspace of \mathbb{R}^{N+P} . We identify \mathbb{R}^{N+P} with the real vector space generated by $\mathcal{S} \cup \mathcal{Y}$ so that $\mathbb{R}^{N+P} \equiv \langle S_1, \dots, S_N, Y_1, \dots, Y_P \rangle$. The species S_i and Y_k are unit vectors with a one in the i -th and $(N+k)$ -th entry, respectively, and all other entries being zero. A vector $v = (\lambda_1, \dots, \lambda_N, \mu_1, \dots, \mu_P)$ is identified with the linear combination of species $\sum_i \lambda_i S_i + \sum_k \mu_k Y_k$.

Consider the *stoichiometric subspace* of \mathbb{R}^{N+P} [3] of a PTM system:

$$\Gamma = \langle S_i + S_j - Y_k \mid (i, j, k) \in I_a \cup I_b \rangle + \langle Y_k - Y_v \mid (k, v) \in I_c \rangle + \langle S_i - S_j \mid (i, j) \in I_d \rangle.$$

If $(\lambda_1, \dots, \lambda_N, \mu_1, \dots, \mu_P) \in \Gamma^\perp$, then $\sum_i \lambda_i \dot{S}_i + \sum_k \mu_k \dot{Y}_k = 0$. The converse might not be true [8]. It follows that any basis $\{\omega^1, \dots, \omega^d\}$ of Γ^\perp provides a set of independent conserved quantities $\sum_{i=1}^N \lambda_i^l S_i + \sum_{k=1}^P \mu_k^l Y_k$ if $\omega^l = (\lambda_1^l, \dots, \lambda_N^l, \mu_1^l, \dots, \mu_P^l)$. Therefore, if *total amounts* $\bar{S}_1, \dots, \bar{S}_d \in \mathbb{R}_+$ are provided, we require the steady state solutions to satisfy:

$$(2.5) \quad \bar{S}_l = \sum_{i=1}^N \lambda_i^l S_i + \sum_{k=1}^P \mu_k^l Y_k \quad l = 1, \dots, d.$$

Total amounts are fixed by the initial concentrations of the species. We say that equations (2.5) are independent if the system has maximal rank, or equivalently, if the corresponding vectors of Γ^\perp are independent.

We introduce the concepts of a cut and a non-interacting graph and show that they provide means to obtain conservation laws.

Definition 2.6. Let a non-empty set $\mathcal{S}_\alpha \subseteq \mathcal{S}$ be given and let the associated set $\mathcal{Y}_\alpha \subseteq \mathcal{Y}$ be the smallest set such that $Y_k \in \mathcal{Y}_\alpha$ if Y_k is 1-linked to some $S_i \in \mathcal{S}_\alpha$ or to $Y_m \in \mathcal{Y}_\alpha$.

- (i) \mathcal{S}_α is *closed* if S_j belongs to \mathcal{S}_α whenever $S_i \in \mathcal{S}_\alpha$ is 1-linked to S_j , and if S_i and S_j interact and are 1-linked to $Y_k \in \mathcal{Y}_\alpha$, then S_i or S_j are in \mathcal{S}_α .
- (ii) \mathcal{S}_α is a *cut* if (a) $S_i, S_j \in \mathcal{S}_\alpha$ do not interact for any i, j , and (b) \mathcal{S}_α is closed.
- (iii) A cut \mathcal{S}_α is *minimal* if it has no proper closed subsets.

Condition (ii) implies that a self-interacting substrate $S \in \mathcal{S}_\alpha$ cannot belong to any cut, that is, $\mathcal{S}_\alpha \cap \mathcal{S}_\alpha = \emptyset$ for any cut \mathcal{S}_α . Note that a closed subset \mathcal{S}' of a cut is also a cut. The union of two disjoint cuts $\mathcal{S}_\alpha, \mathcal{S}'_\alpha$ is a cut if $\mathcal{Y}_\alpha \cap \mathcal{Y}'_\alpha = \emptyset$.

In the PTM system with reactions $S_1 + S_4 \rightleftharpoons Y_2 \rightleftharpoons S_2 + S_4$ and $Y_1 \rightleftharpoons S_2 + S_3$, the set $\{S_1, S_2\}$ is a cut, while $\{S_1, S_3\}$ is not. There are no proper closed subsets of $\{S_1, S_2\}$ and thus the cut is minimal.

Definition 2.7. Let a non-empty set $\mathcal{S}_\alpha \subseteq \mathcal{S}$ be given and let $\mathcal{Y}_\alpha \subseteq \mathcal{Y}$ be as in Definition 2.6. Further, let $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ be the graph with node set $\mathcal{S}_\alpha \cup \mathcal{Y}_\alpha$ and edges between 1-linked nodes. The graph is *non-interacting* if it is connected and \mathcal{S}_α is a cut.

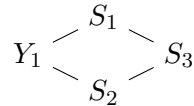
If $\mathcal{S}_\alpha = \mathcal{S}$, then $\mathcal{Y}_\alpha = \mathcal{Y}$. All graphs $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ are naturally subgraphs of $G_{\mathcal{S}, \mathcal{Y}}$. Without proof we state the following:

Lemma 2.8. Let \mathcal{S}_α be a cut and G' be a connected subgraph of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ with node set $\mathcal{S}' \cup \mathcal{Y}'$, $\mathcal{S}' \subseteq \mathcal{S}_\alpha$ and $\mathcal{Y}' \subseteq \mathcal{Y}_\alpha$. The following are equivalent:

- (i) \mathcal{S}' is closed with associated set \mathcal{Y}' .
- (ii) G' is a connected component of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$.
- (iii) G' is non-interacting and contains only species in $\mathcal{S}_\alpha \cup \mathcal{Y}_\alpha$.

If either is the case, then \mathcal{S}' is a minimal cut and $G' = G_{\mathcal{S}', \mathcal{Y}'}$.

Thus, the non-interacting graphs containing substrates only in a cut \mathcal{S}_α are exactly the connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$. All non-interacting graphs contain some node from \mathcal{S} (condition (ii) of a PTM system). However, such a graph might not exist. Consider for example the system with reactions $S_1 \rightleftharpoons S_3$, $S_2 \rightleftharpoons S_3$, $S_1 + S_2 \rightleftharpoons Y_1$. The graph $G_{\mathcal{S}, \mathcal{Y}}$ is



Condition (b) of Definition 2.6(ii) implies that any non-interacting graph must contain all four species, which contradicts condition (a) of the same definition.

Lemma 2.9. *Let H_1, \dots, H_n be the non-interacting graphs of a PTM system, C_l the node set of H_l , $\mathcal{S}_l = \mathcal{S} \cap C_l$ and $\mathcal{Y}_l = \mathcal{Y} \cap C_l$. Then, $\dot{\omega}_l = 0$ for*

$$\omega_l = \sum_{S \in \mathcal{S}_l} S + \sum_{Y \in \mathcal{Y}_l} Y \quad l = 1, \dots, n.$$

That is, H_l corresponds to a conservation law and ω_l is fixed by the initial amounts.

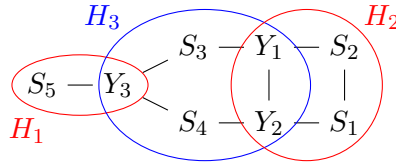
Proof. Substrates in C_l interact only with substrates in $\mathcal{S} \setminus \mathcal{S}_l$ and thus, by definition of \mathcal{Y}_l , if $a_{i,j}^k \neq 0$ or $b_{i,j}^k \neq 0$ for $i \neq j$ then: (a) if S_i (resp. S_j) is in \mathcal{S}_l , then S_j (resp. S_i) belongs to $\mathcal{S} \setminus \mathcal{S}_l$, and $Y_k \in \mathcal{Y}_l$; (b) if $Y_k \in \mathcal{Y}_l$, then either S_i or S_j , but not both, belongs to \mathcal{S}_l . If $c_{v,k} \neq 0$ or $c_{k,v} \neq 0$, then Y_k, Y_v belong to the same non-interacting graph (if any); if $d_{i,j} \neq 0$ or $d_{j,i} \neq 0$, then S_i, S_j belong to the same non-interacting graph (if any). Since $\mathcal{S}_l \cap \mathcal{S}_o = \emptyset$ for $Y_k \in \mathcal{Y}_l$ and $S_i \in \mathcal{S}_l$ we have:

$$\begin{aligned} \dot{Y}_k &= \sum_{i|S_i \in \mathcal{S}_l} \sum_{j|S_j \in \mathcal{S} \setminus \mathcal{S}_l} (a_{i,j}^k S_i S_j - b_{i,j}^k Y_k) + \sum_{v|Y_v \in \mathcal{Y}_l} (c_{v,k} Y_v - c_{k,v} Y_k) \\ \dot{S}_i &= \sum_{k|Y_k \in \mathcal{Y}_l} \sum_{j|S_j \in \mathcal{S} \setminus \mathcal{S}_l} (-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j|S_j \in \mathcal{S}_l} (d_{j,i} S_j - d_{i,j} S_i). \end{aligned}$$

It follows that $\sum_{k|Y_k \in \mathcal{Y}_l} \sum_{v|Y_v \in \mathcal{Y}_l} (c_{v,k} Y_v - c_{k,v} Y_k) = 0$ and $\sum_{i|S_i \in \mathcal{S}_l} \sum_{j|S_j \in \mathcal{S}_l} (d_{j,i} S_j - d_{i,j} S_i) = 0$. Similarly, the remaining terms in $\dot{\omega}_l$ cancel. Thus, $\dot{\omega}_l = 0$. \square

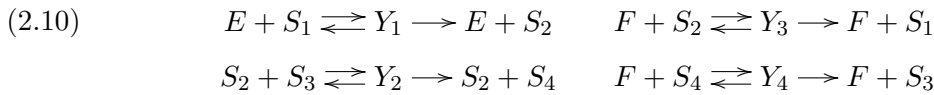
Thus, each non-interacting graph gives rise to a conserved amount. If each non-interacting graph contains a species that only belongs to that specific graph, then the ω_l 's are independent. In particular, conservation laws derived from the connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ for some cut \mathcal{S}_α are independent. In general, the set of conservation laws found from Lemma 2.9 can be reduced to a set of independent conservation laws.

In Example (2.1), the graph $G_{\mathcal{S}, \mathcal{Y}}$ is



The non-interacting graphs H_1, H_2, H_3 are colored. If total amounts $\bar{S}_1, \bar{S}_2, \bar{S}_3$ are provided then the steady state solutions must satisfy: $\bar{S}_1 = S_5 + Y_3$, $\bar{S}_2 = S_1 + S_2 + Y_1 + Y_2$, and $\bar{S}_3 = S_3 + S_4 + Y_1 + Y_2 + Y_3$. These conserved total amounts are easily verified by differentiation using (2.4).

Consider a two-layer cascade of modification cycles that share the demodification enzyme F in each layer. The reaction system consists of $\mathcal{S} = \{E, F, S_1, S_2, S_3, S_4\}$, $\mathcal{Y} = \{Y_1, Y_2, Y_3, Y_4\}$ and the reactions



The subsets $\mathcal{S}_\alpha = \{E, S_3, S_4\}, \{E, F\}, \{S_1, S_2\}$ are examples of maximal cuts (they cannot be extended to larger cuts by including more substrates). The graph $G_{\mathcal{S}, \mathcal{Y}}$ is

In the following we make use of Lemma 2.8 without further reference. Let \mathcal{S}_α be a cut with associated set \mathcal{Y}_α . Define $\mathcal{S}_\alpha^\complement = \mathcal{S} \setminus \mathcal{S}_\alpha$ and $\mathcal{Y}_\alpha^\complement = \mathcal{Y} \setminus \mathcal{Y}_\alpha$, and let N_α , P_α (resp. N_α^\complement , P_α^\complement) be the cardinality of \mathcal{S}_α , \mathcal{Y}_α (resp. $\mathcal{S}_\alpha^\complement$, $\mathcal{Y}_\alpha^\complement$). Extend the set of conservation laws derived from the connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ to a maximal set of n independent conservation

laws derived from *other* non-interacting graphs (thus containing species in $\mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c$). Let $n_\alpha^c = n - n_\alpha$, where n_α is the number of connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$.

Lemma 2.13. *Let \mathcal{S}_α be a cut and keep the notation introduced above. Then, we have that $\dim(\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma) \leq N_\alpha^c + P_\alpha^c - n_\alpha^c$ and $\dim \Gamma^\perp = n$ if and only if*

$$\dim(\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma) = N_\alpha^c + P_\alpha^c - n_\alpha^c.$$

Proof. Without loss of generality we can assume that $\mathcal{Y}_\alpha = \{Y_1, \dots, Y_{P_\alpha}\}$ and $\mathcal{S}_\alpha = \{S_1, \dots, S_{N_\alpha}\}$. Identify \mathbb{R}^{N+P} with $\mathbb{R}^{N_\alpha} \times \mathbb{R}^{P_\alpha} \times \mathbb{R}^{N_\alpha^c} \times \mathbb{R}^{P_\alpha^c}$ and let

$$\Gamma_\alpha = \langle A - B \mid \text{for each edge } A - B \text{ in } G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha} \rangle.$$

The space Γ_α^\perp is generated by the vectors which are sums of species in each connected component of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ and hence $\dim \Gamma_\alpha^\perp = n_\alpha$. We have $\dim \Gamma^\perp \geq n = n_\alpha + n_\alpha^c$ and we want to determine when equality holds. Equivalently, we want to see when $\dim \Gamma = N + P - n$. If this is not the case, then $\dim \Gamma < N + P - n$. Note that $N = N_\alpha + N_\alpha^c$ and $P = P_\alpha + P_\alpha^c$.

Note that $\dim \Gamma_\alpha = N_\alpha + P_\alpha - n_\alpha$. Let $\pi: \mathbb{R}^{N+P} \rightarrow \mathbb{R}^{N_\alpha+P_\alpha}$ denote the projection onto the first $N_\alpha + P_\alpha$ coordinates and $\pi_\alpha: \Gamma \rightarrow \Gamma_\alpha$ its restriction to Γ (π_α a surjective map). Then, $\dim \Gamma = \dim \Gamma_\alpha + \dim \ker \pi_\alpha$ and so $\dim \ker \pi_\alpha \leq N_\alpha^c + P_\alpha^c - n_\alpha^c$. Further, $\dim \Gamma^\perp = n$ if and only if $\dim \ker \pi_\alpha = N_\alpha^c + P_\alpha^c - n_\alpha^c$. Finally, note that $\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma = \ker \pi_\alpha$. Indeed, let $i: \Gamma \hookrightarrow \mathbb{R}^{N+P}$ and $i_\alpha: \Gamma_\alpha \hookrightarrow \mathbb{R}^{N_\alpha+P_\alpha}$ denote the natural inclusions. We have that $i_\alpha \circ \pi_\alpha = \pi \circ i$. The kernel of π is clearly $\mathbb{R}^{N_\alpha^c+P_\alpha^c} = \langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle$ from where it follows that the kernel of π_α is $\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma$.

Therefore, $\dim(\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma) = \dim \ker \pi_\alpha = N_\alpha^c + P_\alpha^c - n_\alpha^c$ if and only if $\dim \Gamma^\perp = n$ and the lemma is proved. \square

As each non-interacting graph corresponds to a minimal cut, the lemma above provides a condition for when all conservation laws are recovered from cuts.

Remark. An easy way to construct elements of $\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma$ is by considering:

- (i) Vectors $S_i - S_j$ for any pair $S_i, S_j \in \mathcal{S}_\alpha^c$ for which there exists a chain of reactions $S_m + S_i - A_1 - \dots - A_r - S_m + S_j$ for some $S_m, A_u \in \mathcal{C}$ and $-$ is \leftarrow or \rightarrow .
- (ii) Vectors $S_i + S_j - Y_k$, $S_i - S_j$ or $Y_k - Y_v$ corresponding to reactions with $S_i, S_j \in \mathcal{S}_\alpha^c$ and $Y_k, Y_v \in \mathcal{Y}_\alpha^c$.

If we can construct $N_\alpha^c + P_\alpha^c - n_\alpha^c$ independent elements of $\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma$ of the previous type, then the previous lemma holds.

In Example (2.1) consider the cut $\mathcal{S}_\alpha = \{S_1, S_2, S_5\}$ with $\mathcal{S}_\alpha^c = \{S_3, S_4\}$ and the given conservation laws ($n = 3$). We have $N_\alpha^c = 2$ and $n_\alpha^c = 1$. Further, $\mathcal{Y}_\alpha = \mathcal{Y}$ so that $P_\alpha^c = 0$. The element $S_3 - S_4 = (S_3 + S_5 - Y_3) - (S_4 + S_5 - Y_3)$ belongs to $\langle \mathcal{S}_\alpha^c \rangle \cap \Gamma$. In addition, $N_\alpha^c + P_\alpha^c - n_\alpha^c = 1$ and thus $\dim(\langle \mathcal{S}_\alpha^c \rangle \cap \Gamma) = N_\alpha^c + P_\alpha^c - n_\alpha^c$, implying that all conservation laws are found from non-interacting graphs.

In Example (2.10), consider the cut $\mathcal{S}_\alpha = \{E, S_3, S_4\}$ with $\mathcal{S}_\alpha^c = \{F, S_1, S_2\}$ and $N_\alpha^c = 3$. In this case, $\mathcal{Y}_\alpha = \{Y_1, Y_2, Y_4\}$, $\mathcal{Y}_\alpha^c = \{Y_3\}$ and so $P_\alpha^c = 1$. Two of the four conservation laws involve elements in $\mathcal{S}_\alpha \cup \mathcal{Y}_\alpha$ only and hence $n_\alpha = n_\alpha^c = 2$. Further, $N_\alpha^c + P_\alpha^c - n_\alpha^c = 2$. The two independent vectors $F + S_2 - Y_3$ and $F + S_1 - Y_3$ belong to $\langle \mathcal{S}_\alpha^c \cup \mathcal{Y}_\alpha^c \rangle \cap \Gamma$. Thus, the graphical procedure provides all conservation laws.

In Example (2.11), consider the cut $\mathcal{S}_\alpha = \{S_1, S_3\}$ with $\mathcal{S}_\alpha^c = \{S_2, S_4\}$ so that $N_\alpha^c = 2$, $P_\alpha^c = 0$. There is only one conservation law in $\mathcal{S}_\alpha \cup \mathcal{Y}_\alpha$, $S_3 + S_1 + Y_1 + Y_2$, and since $n = 3$, then $n_\alpha^c = 2$. It follows that $N_\alpha^c + P_\alpha^c - n_\alpha^c = 0$, and we are guaranteed that the dimension of $\langle S_2, S_4 \rangle \cap \Gamma$ is zero.

In Example (2.12), consider the cut $\mathcal{S}_\alpha = \{S_1, S_2\}$ with $\mathcal{S}_\alpha^c = \{S_3, S_4\}$ and $N_\alpha^c = 2$, $P_\alpha^c = 0$. We have $n_\alpha^c = 1$ and $N_\alpha^c + P_\alpha^c - n_\alpha^c = 1$. However, $\langle S_3, S_4 \rangle \cap \Gamma$ has dimension zero and thus not all conservation laws arise from non-interacting graphs.

3. VARIABLE ELIMINATION

In this section we show that the intermediate complexes can always be eliminated and expressed as polynomials in the substrates with coefficients in $\mathbb{R}(\text{Con})$ (Section 3.1). After choosing a cut \mathcal{S}_α , the substrates in \mathcal{S}_α can be expressed in terms of those in $\mathcal{S}_\alpha^c = \mathcal{S} \setminus \mathcal{S}_\alpha$ (Section 3.3).

3.1. Elimination of intermediate complexes. Consider the system $\dot{Y}_i = 0$ in (2.2) as a linear system of P polynomial equations with coefficients in $\mathbb{R}[\text{Con} \cup \mathcal{S}]$ and P variables Y_1, \dots, Y_P . If the system has maximal rank, then there is a unique solution in $\mathbb{R}(\text{Con} \cup \mathcal{S})$.

Specifically, we have a linear system $AY = z$ where $Y = (Y_1, \dots, Y_P)^t$ and $A = \{\lambda_{k,v}\}$ is a $P \times P$ matrix with coefficients in $\mathbb{R}[\text{Con}]$,

$$\lambda_{k,v} = \begin{cases} -c_{v,k} & \text{if } k \neq v \\ \sum_{j=1}^N \sum_{i=1}^j b_{i,j}^k + \sum_{u=1}^P c_{k,u} & \text{if } k = v. \end{cases}$$

The independent term $z = (z_1, \dots, z_P)^t$ is in $\mathbb{R}[\text{Con} \cup \mathcal{S}]$: $z_k = \sum_{i \leq j} a_{i,j}^k S_i S_j$.

Assume that A has maximal rank P in $\mathbb{R}(\text{Con})$. Then, using Cramer's rule to solve linear systems of equations, we obtain that $Y_k = \rho_k / \rho$ with $\rho = \det(A) \neq 0$ and ρ_k the determinant of A with the k -th column substituted by z . Since the determinant is a homogeneous polynomial in the entries of the matrix, it follows that $\rho \in \mathbb{R}[\text{Con}]$ and $\rho_k \in \mathbb{R}[\text{Con} \cup \mathcal{S}]$. Therefore,

$$Y_k = \sum_{i \leq j} \mu_{i,j}^k S_i S_j$$

with $\mu_{i,j}^k \in \mathbb{R}(\text{Con})$ and thus Y_k is a polynomial in $\mathbb{R}(\text{Con})[\mathcal{S}]$. If both ρ, ρ^k are S -positive elements of $\mathbb{R}[\text{Con}]$ and $\mathbb{R}[\text{Con} \cup \mathcal{S}]$, respectively, then for positive rate constants and non-negative values of S_i , the steady state value of Y_k is non-negative as well. S -positivity of ρ, ρ^k is proven in the next section using the *Matrix-Tree theorem* [19]. Some basic concepts from graph theory are required.

Graphs and the Matrix-Tree theorem. Given a directed graph G , a *spanning tree* τ is a directed subgraph with the same node set as G and such that the corresponding undirected graph is connected and acyclic. There is a unique undirected path between any two nodes in a spanning tree [5]. A spanning tree τ is said to be *rooted* at a node v if the unique path between any node w and v is directed from w to v . It follows that v is the only node with no out-edges, that is, there is no edge of the form $v \rightarrow w$ in τ . In addition, there cannot be a node with two out-edges in τ . The graph G is *strongly connected* if for any pair of nodes v, w there is a directed path from v to w . Any directed path from v to

w in a strongly connected graph can be extended to a spanning tree rooted at w . Some general references for graph theory are [5] and [10].

If G is labeled, then τ inherits a labeling from G and we define

$$\pi(\tau) = \prod_{x \xrightarrow{a} y \in \tau} a.$$

Assume that G has no self-loops. Order the node set $\{v_1, \dots, v_n\}$ of G and denote by $a_{i,j}$ the label of the edge $v_i \rightarrow v_j$. We set $a_{i,j} = 0$ if there is no edge from v_i to v_j (thus $a_{i,i} = 0$). Let $\mathcal{L}(G) = \{\alpha_{i,j}\}$ be the *Laplacian* of G , that is the matrix with

$$\alpha_{i,j} = \begin{cases} a_{j,i} & \text{if } i \neq j \\ -\sum_{k=1}^n a_{i,k} & \text{if } i = j, \end{cases}$$

such that the column sums are zero. For each node v_j , let $\Theta(v_j)$ be the set of spanning trees of G rooted at v_j . Then, the Matrix-Tree theorem states that the maximal minor $\mathcal{L}(G)_{(ij)}$ (the determinant of the minor obtained by removing the i -th row and the j -th column of $\mathcal{L}(G)$) is:

$$\mathcal{L}(G)_{(ij)} = (-1)^{n-1+i+j} \sum_{\tau \in \Theta(v_j)} \pi(\tau).$$

Note that for notational simplicity we have defined the Laplacian as the transpose of how it is usually defined and the Matrix-Tree theorem has been adapted consequently.

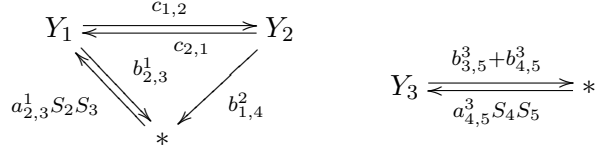
In our case, the matrix A is not a Laplacian, since the column sums $\sum_{j=1}^N \sum_{i=1}^j b_{i,j}^k$ are not zero. However, A can be extended such that its determinant is a maximal minor of a Laplacian.

3.2. Decomposition of the system. Let $G_{\mathcal{Y}}$ be the directed graph with node set \mathcal{Y} and a directed edge $Y_k \rightarrow Y_v$ if $(k, v) \in I_c$. The node sets of the connected components of $G_{\mathcal{Y}}$ determine a partition of \mathcal{Y} : $\mathcal{Y} = \mathcal{Y}_1 \cup \dots \cup \mathcal{Y}_s$. Let P_l be the cardinality of \mathcal{Y}_l and rename the intermediate complexes such that $\mathcal{Y}_l = \{Y_{P_1+\dots+P_{l-1}+1}, \dots, Y_{P_1+\dots+P_l}\}$.

If $Y_k \in \mathcal{Y}_l$ for some l , then $c_{k,v} = c_{v,k} = 0$ for any v such that $Y_v \notin \mathcal{Y}_l$. It follows, that A is a block diagonal matrix $\text{diag}(A_1, \dots, A_s)$ with A_l being a $P_l \times P_l$ matrix. Solving $AY = z$ is thus equivalent to solving s “smaller” systems with matrices A_l . Further, A has maximal rank P if and only if A_l has maximal rank P_l for all l .

Consider the connected component $G_{\mathcal{Y}_l}$ corresponding to \mathcal{Y}_l . We construct an extended labeled directed graph $\hat{G}_{\mathcal{Y}_l}$ with node set $\mathcal{Y}_l \cup \{*\}$. For convenience we order the nodes such that $Y_{P_1+\dots+P_{l-1}+k}$ is the k -th node and $*$ the $(P_l + 1)$ -th node. Let $b^k = \sum_{i \leq j} b_{i,j}^k$ and $a^k = \sum_{i \leq j} a_{i,j}^k$. The graph $\hat{G}_{\mathcal{Y}_l}$ has the following labeled directed edges: $Y_k \xrightarrow{c_{k,v}} Y_v$ if $(k, v) \in I_c$, $Y_k \xrightarrow{b^k} *$ if $b^k \neq 0$, and $* \xrightarrow{a^k} Y_k$ if $a^k \neq 0$.

In Example (2.1), the graph $G_{\mathcal{Y}}$ has two connected components $\mathcal{Y}_1 = Y_1 \rightleftharpoons Y_2$ and $\mathcal{Y}_2 = Y_3$. The graphs $\hat{G}_{\mathcal{Y}_1}$ and $\hat{G}_{\mathcal{Y}_2}$ are



Let $\mathcal{L} = \{\alpha_{k,v}\}$ be the Laplacian of $\widehat{G}_{\mathcal{Y}_l}$. If $k, v \leq P_l$, then $\alpha_{k,v} = -\lambda_{k,v}$. The entries of the last row are $\alpha_{P_l+1,k} = b^k$ for $k \leq P_l$ and the entries of the last column are $\alpha_{k,P_l+1} = a^k$ ($= z_k$) for $k \leq P_l$. We conclude that the $(P_l + 1, P_l + 1)$ principal minor of \mathcal{L} is exactly $-A_l$ and thus, by the Matrix-Tree theorem, we have

$$\det(A_l) = (-1)^{P_l} \mathcal{L}_{(P_l+1, P_l+1)} = \sum_{\tau \in \Theta(*)} \pi(\tau).$$

Assumption (ii) of a PTM system ensures that each Y_k ultimately reacts to some $S_i + S_j$ via \mathcal{Y} , and hence there exists at least one spanning tree rooted at $*$. Thus, $\det(A_l) \neq 0$ and $\det(A_l)$ is an S-positive element of $\mathbb{R}[\text{Con}]$.

By the definition of ρ_k and the Matrix-Tree theorem,

$$\rho_k = (-1)^{k+1} \mathcal{L}_{(P_l+1, k)} = \sum_{\tau \in \Theta(Y_{P_l+1+k})} \pi(\tau),$$

and hence ρ_k is either zero or an S-positive element of $\mathbb{R}[\text{Con} \cup \mathcal{S}]$.

If there exists at least one spanning tree rooted at $v_k = Y_{P_l+1+k}$, then $\rho_k \neq 0$. A necessary condition for this to happen is the existence of at least one in-edge to v_k . Otherwise the concentration at steady state of v_k is zero, which is expected if v_k is only consumed and never produced. Similarly, if there is no reaction of the form $S_i + S_j \rightarrow Y_{P_l+1+m}$ for any m (that is, a directed edge $* \rightarrow v_m$), then $\rho_k = 0$ for all k .

The term ρ_k is a homogeneous polynomial of degree 2 in \mathcal{S} with coefficients in $\mathbb{R}[\text{Con}]$, because any spanning tree rooted at a node v_k has exactly one edge of the form $* \rightarrow v_m$ for some m . Further, a monomial $S_i S_j$ appears in ρ_k only if $S_i + S_j$ ultimately reacts to v_k via \mathcal{Y}_l . If $\widehat{G}_{\mathcal{Y}_l}$ is strongly connected, then this condition is both sufficient and necessary. Indeed, if $S_i + S_j$ ultimately reacts to v_k via \mathcal{Y}_l , then there is a spanning tree rooted at v_k containing this path.

The next proposition summarizes the discussion above:

Proposition 3.1. *Consider a PTM system with intermediate complexes \mathcal{Y} and substrates \mathcal{S} . Then, $\dot{Y}_k = 0$ for all k , if and only if*

$$(3.2) \quad Y_k = \sum_{i \leq j} \mu_{i,j}^k S_i S_j$$

with $\mu_{i,j}^k \in \mathbb{R}(\text{Con})$ being either zero or S-positive. Further:

- (i) If $S_i + S_j$ does not ultimately react to Y_k via \mathcal{Y} , then $\mu_{i,j}^k = 0$.
- (ii) If $\widehat{G}_{\mathcal{Y}_l}$ is strongly connected and $Y_k \in \mathcal{Y}_l$, then $\mu_{i,j}^k \neq 0$ if and only if $S_i + S_j$ ultimately reacts to Y_k via \mathcal{Y}_l .
- (iii) $\widehat{G}_{\mathcal{Y}_l}$ is strongly connected if and only if in (3.2), Y_k is a non-zero polynomial in $\mathbb{R}(\text{Con})[\mathcal{S}]$ for all $Y_k \in \mathcal{Y}_l$.

Remark. The condition that $\widehat{G}_{\mathcal{Y}_l}$ is strongly connected is biochemically reasonable: The intermediate complexes are not the initial or final products of the system and should eventually be broken up into parts.

In Example (2.1), the graph $\widehat{G}_{\mathcal{Y}_1}$ has three spanning trees rooted at $*$ so that $\det(A_1) = b_{1,4}^2 c_{1,2} + b_{2,3}^1 c_{2,1} + b_{1,4}^2 b_{2,3}^1$. There is one spanning tree rooted at Y_2 , giving $\rho_2 = c_{1,2} a_{2,3}^1 S_2 S_3$, and two spanning trees rooted at Y_1 , giving $\rho_1 = (b_{1,4}^2 + c_{2,1}) a_{2,3}^1 S_2 S_3$. The graph $\widehat{G}_{\mathcal{Y}_2}$ has one spanning tree rooted at $*$ so that $\det(A_2) = b_{3,5}^3 + b_{4,5}^3$, and one spanning tree rooted at Y_3 , giving $\rho_3 = a_{4,5}^3 S_4 S_5$. Thus:

$$Y_1 = \mu_{2,3}^1 S_2 S_3, \quad Y_2 = \mu_{2,3}^2 S_2 S_3, \quad Y_3 = \mu_{4,5}^3 S_4 S_5$$

$$\text{with } \mu_{2,3}^1 = \frac{(b_{1,4}^2 + c_{2,1}) a_{2,3}^1}{\det(A_1)}, \mu_{2,3}^2 = \frac{c_{1,2} a_{2,3}^1}{\det(A_1)}, \text{ and } \mu_{4,5}^3 = \frac{a_{4,5}^3}{\det(A_2)}.$$

Lemma 3.3. *Let $\widehat{G}_{\mathcal{Y}} = \cup_l \widehat{G}_{\mathcal{Y}_l}$. The graphs $\widehat{G}_{\mathcal{Y}_l}$, $l = 1, \dots, s$, are strongly connected if and only if the graph $\widehat{G}_{\mathcal{Y}}$ is.*

Proof. Assume that the graphs $\widehat{G}_{\mathcal{Y}_l}$ are strongly connected. Then, for any $v \in \widehat{G}_{\mathcal{Y}_l}$ and $w \in \widehat{G}_{\mathcal{Y}_j}$, there are directed paths $v \rightarrow *$ in $\widehat{G}_{\mathcal{Y}_l}$ and $* \rightarrow w$ in $\widehat{G}_{\mathcal{Y}_j}$, which by composition give a directed path between v and w .

For the reverse implication, let v, w be two elements of \mathcal{Y}_l . Since $\widehat{G}_{\mathcal{Y}}$ is strongly connected, there exists a directed path $\alpha: v \rightarrow w$ in $\widehat{G}_{\mathcal{Y}}$. We can assume that $v, w \neq *$. A path connecting an intermediate complex in \mathcal{Y}_l to one in \mathcal{Y}_j for $j \neq l$ must pass through $*$. If a path α goes through $\tilde{v} \in \mathcal{Y}_j$, for $j \neq l$, then it must go through $*$, first in and then out, potentially many times until it goes back to \mathcal{Y}_l and to w . Therefore, α has the form $v \xrightarrow{\alpha_1} * \xrightarrow{\beta} * \xrightarrow{\alpha_2} w$ with α_1 and α_2 being paths in $\widehat{G}_{\mathcal{Y}_l}$. It follows that the path $v \xrightarrow{\alpha_1} * \xrightarrow{\alpha_2} w$ is a directed path from v to w in $\widehat{G}_{\mathcal{Y}_l}$. \square

3.3. Elimination of substrates. Equation (3.2) shows that at steady state the intermediate complexes are given as zero or S-positive rational functions in the substrates and the rate constants. Insertion of (3.2) into the (time dependent) differential equations for the substrates is the procedure known as the *quasi-steady state* assumption. The rationale is that intermediate complexes tend to reach steady state much faster than substrates and thus some variables in the dynamical system can be eliminated. We have shown here that PTM systems “mathematically” enable this simplification although justification is required in concrete examples.

We now use the steady state equation (2.3) to further eliminate some of the substrates in terms of others. Recall equation (2.3), that is $\dot{S}_i = 0$,

$$(3.4) \quad 0 = \sum_{j=1}^N \sum_{k=1}^P \epsilon_{i,j} (-a_{i,j}^k S_i S_j + b_{i,j}^k Y_k) + \sum_{j=1}^N (d_{j,i} S_j - d_{i,j} S_i)$$

for $i = 1, \dots, N$. After substitution of the values for Y_k , we have

$$(3.5) \quad 0 = \sum_{u=1}^N \sum_{k=1}^P \sum_{j \leq t} \epsilon_{i,j} b_{i,u}^k \mu_{j,t}^k S_j S_t - \sum_{j=1}^N \sum_{k=1}^P \epsilon_{i,j} a_{i,j}^k S_i S_j + \sum_{j=1}^N (d_{j,i} S_j - d_{i,j} S_i),$$

These equations are quadratic in \mathcal{S} . To proceed with linear elimination it is necessary to decide which variables are to be eliminated and which will be taken as part of the coefficient field. Since a monomial $S_i S_j$ appears only if S_i and S_j interact, we can proceed as long as \mathcal{S} can be partitioned in an appropriate way.

Lemma 3.6. *Assume that there is a cut \mathcal{S}_α with associated set \mathcal{Y}_α (Definition 2.6). Then, (3.5) for the substrates in \mathcal{S}_α is a homogeneous linear system of equations in the substrates \mathcal{S}_α and with coefficients in $\mathbb{R}(\text{Con} \cup \mathcal{S}_\alpha^c)$.*

Proof. For the three sums in (3.5) we make the following observations: If $S_i \in \mathcal{S}_\alpha$ and $d_{j,i} \neq 0$, then also $S_j \in \mathcal{S}_\alpha$. If $S_i \in \mathcal{S}_\alpha$ and $a_{i,j}^k \neq 0$, then $S_j \notin \mathcal{S}_\alpha$, otherwise S_i and S_j would interact. Finally, if $\mu_{j,t}^k \neq 0$, then according to Proposition 3.1, $S_j + S_t$ ultimately reacts to Y_k via \mathcal{Y}_α . Hence, since \mathcal{S}_α is a cut, one of S_j and S_t (but not both) belongs to \mathcal{S}_α . Thus (3.5) for the substrates in \mathcal{S}_α is a homogeneous linear system of equations in the species in \mathcal{S}_α . \square

Assume that there exists a cut \mathcal{S}_α and that $\mathcal{S}_\alpha = \{S_1, \dots, S_{N_\alpha}\}$. It follows that for $S_i \in \mathcal{S}_\alpha$ the equations in (3.5) form an $N_\alpha \times N_\alpha$ homogeneous linear system of equations with variables \mathcal{S}_α and coefficients in $\mathbb{R}(\text{Con} \cup \mathcal{S}_\alpha^c)$. Further, $\epsilon_{i,j} = 1$ if $S_i \in \mathcal{S}_\alpha$.

Let B be the matrix with entries $\tilde{b}_{i,j}$ for $i \neq j$ and $\tilde{b}_{i,i} - \tilde{a}_i$ for $i = j$, where

$$\tilde{a}_i = \sum_{j=1}^{N_1} d_{i,j} + \sum_{j=N_1+1}^N \sum_{k=1}^P a_{i,j}^k S_j, \quad \tilde{b}_{i,j} = d_{j,i} + \sum_{t=N_1+1}^N \sum_{k=1}^P b_{i,j,t}^k \mu_{j,t}^k S_t, \quad b_i^k = \sum_{u=N_1+1}^N b_{i,u}^k,$$

so that (3.5) becomes

$$(3.7) \quad 0 = \sum_{j=1, j \neq i}^{N_1} \tilde{b}_{i,j} S_j + (\tilde{b}_{i,i} - \tilde{a}_i) S_i.$$

Consider Example (2.1) and the cut $\mathcal{S}_\alpha = \{S_1, S_2, S_5\}$. Then the equations (3.5) are $0 = -d_{1,2} S_1 + b_{1,4}^1 \mu_{2,3}^1 S_2 S_3$ and $0 = d_{1,2} S_1 - a_{2,3}^1 S_2 S_3 + b_{2,3}^1 \mu_{2,3}^1 S_2 S_3$, corresponding to $\dot{S}_1 = 0$ and $\dot{S}_2 = 0$, respectively. The equation $\dot{S}_5 = 0$ is trivial because of the conservation law $\dot{Y}_3 + \dot{S}_5 = 0$. Further, we have $\tilde{a}_1 = d_{1,2}$, $\tilde{a}_2 = a_{2,3}^1 S_3$, $\tilde{b}_{1,2} = b_{1,4}^1 \mu_{2,3}^1 S_3$, $\tilde{b}_{2,1} = d_{1,2}$, $\tilde{b}_{2,2} = b_{2,3}^1 \mu_{2,3}^1 S_3$, while the rest of the coefficients are zero.

Lemma 2.8 ensures that there is a conservation law for each connected component of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$. Let $C_1^\alpha, \dots, C_{n_\alpha}^\alpha$ be the node sets of the connected components and define $\mathcal{S}_{\alpha,l} = \mathcal{S}_\alpha \cap C_l^\alpha$ and $\mathcal{Y}_{\alpha,l} = \mathcal{Y}_\alpha \cap C_l^\alpha$ so that $\sum_{S_i \in \mathcal{S}_{\alpha,l}} \dot{S}_i + \sum_{Y_k \in \mathcal{Y}_{\alpha,l}} \dot{Y}_k = 0$ for $l = 1, \dots, n_\alpha$, are conservation laws. Imposing only that the intermediate complexes are at steady state, that is $\dot{Y}_k = 0$ for all k , we obtain

$$(3.8) \quad \sum_{S_i \in \mathcal{S}_{\alpha,l}} \dot{S}_i = 0, \quad l = 1, \dots, n_\alpha.$$

It follows that the column sums of the matrix B restricted to the rows corresponding to the substrates in $\mathcal{S}_{\alpha,l}$ are all zero. Consequently, the matrix B has rank at most $N_\alpha - n_\alpha$.

Let $G_{\mathcal{Y}_\alpha}$ be $G_{\mathcal{Y}}$ restricted to the nodes \mathcal{Y}_α . It follows from the definition of \mathcal{Y}_α (Definition 2.6) that $G_{\mathcal{Y}_\alpha}$ is a union of connected components of $G_{\mathcal{Y}}$. Define $\widehat{G}_{\mathcal{Y}_\alpha}$ similarly (cf. Lemma 3.3). Let $N_{\alpha,l}$ be the cardinality of $\mathcal{S}_{\alpha,l}$.

Lemma 3.9. *After reordering the substrates in \mathcal{S}_α , B is a block diagonal matrix, namely $\text{diag}(B_1, \dots, B_{n_\alpha})$, where B_l is an $N_{\alpha,l} \times N_{\alpha,l}$ matrix. Further, if $\widetilde{b}_{i,j} \neq 0$ then there is a reaction $S_j \rightarrow S_i$ or there exist $S_u, S_t \in \mathcal{S}_\alpha^c$, so that $S_j + S_t$ ultimately reacts to $S_i + S_u$ via \mathcal{Y}_α . If in addition $\widehat{G}_{\mathcal{Y}_\alpha}$ is strongly connected, then the reverse is true.*

Proof. It follows from Lemma 3.3 and Proposition 3.1(i) that if $\mu_{j,t}^k \neq 0$ then $S_j + S_t$ ultimately reacts to Y_k . By definition, $b_i^k \neq 0$ if and only if there exists a reaction $Y_k \rightarrow S_i + S_u$ for some $S_u \in \mathcal{S}_\alpha^c$. We have $\widetilde{b}_{i,j} \neq 0$ if and only if $d_{j,i} \neq 0$ or $b_i^k \mu_{j,t}^k \neq 0$ for some k and t , and hence either there is a reaction $S_j \rightarrow S_i$ or there exist $S_u, S_t \in \mathcal{S}_\alpha^c$, so that $S_j + S_t$ ultimately reacts to $S_i + S_u$ via \mathcal{Y}_α . If $\widehat{G}_{\mathcal{Y}_\alpha}$ is strongly connected then by Proposition 3.1(ii) the existence of these reactions is a sufficient condition. It follows, after reordering of the species in \mathcal{S}_α , that B is a block diagonal matrix with blocks given by the species in each connected component of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$. Indeed, if S_i, S_j are in different components, then $\widetilde{b}_{i,j} = \widetilde{b}_{j,i} = 0$. \square

It follows from the lemma that a necessary condition for $\widetilde{b}_{i,j} \neq 0$ is that S_i can be “produced” from S_j . We restrict the study to the case where $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ is connected and note that the results apply to every connected component individually. However, the propositions to be derived below are stated in full generality, that is, without the assumption that $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ is connected.

Using (3.8), the column sums of B are zero. Thus, B is the Laplacian of a labeled directed graph $G_{\mathcal{S}_\alpha}$ with node set \mathcal{S}_α and an edge from S_j to S_i whenever $\widetilde{b}_{i,j} \neq 0$, $i \neq j$. Note that $\widetilde{b}_{i,j} \in \mathbb{R}(\text{Con})[\mathcal{S}_\alpha^c]$ is S-positive.

Since $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ is connected, then so is $G_{\mathcal{S}_\alpha}$. In general, two species S_i, S_j belong to the same connected component of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ if and only if they belong to the same connected component of $G_{\mathcal{S}_\alpha}$. We will use this fact repeatedly in what follows.

By the Matrix-Tree theorem, the principal minors $B_{(i,j)}$ of $B = \mathcal{L}(G_{\mathcal{S}_\alpha})$ are

$$B_{(i,j)} = (-1)^{N_\alpha - 1 + i + j} \sum_{\tau \in \Theta(S_j)} \pi(\tau).$$

Thus, B has rank $N_\alpha - 1$ if and only if there exists at least one spanning tree in $G_{\mathcal{S}_\alpha}$ rooted at some S_j with $j \in \{1, \dots, N_\alpha\}$. For a general PTM system with a selected cut \mathcal{S}_α , we obtain the following proposition:

Proposition 3.10. *The non-interacting graphs provide all conservation laws involving only the substrates $\mathcal{S}_{\alpha,l}$ if and only if $G_{\mathcal{S}_{\alpha,l}}$ has at least one rooted spanning tree for all l .*

Proof. The non-interacting graphs provide all conservation laws involving only $\mathcal{S}_{\alpha,l}$ if and only if all conservation laws are multiples of $\sum_{S_i \in \mathcal{S}_{\alpha,l}} S_i + \sum_{Y_k \in \mathcal{Y}_{\alpha,l}} Y_k = 0$, which is the case if and only if the rank of B_l is $N_{\alpha,l} - 1$. As stated above this is equivalent to the existence of a rooted spanning tree in $G_{\mathcal{S}_{\alpha,l}}$. \square

Remark. In particular, the lemma holds if $G_{\mathcal{S}_\alpha}$ is strongly connected. If $\widehat{G}_{\mathcal{Y}_\alpha}$ is strongly connected, then to check that $G_{\mathcal{S}_\alpha}$ is strongly connected we do not need to calculate the labels of $G_{\mathcal{S}_\alpha}$. Whether there is an edge or not between two nodes follows from the set of reactions, cf. Lemma 3.9.

For simplicity we assume that there exists a spanning tree rooted at S_1 . Then, the variables S_2, \dots, S_{N_α} can be solved in the coefficient field $\mathbb{R}(\text{Con} \cup \mathcal{S}_\alpha^c \cup \{S_1\})$. In particular, using Cramer's rule and the Matrix-Tree theorem, we obtain

$$(3.11) \quad S_j = \frac{(-1)^{j+1} B_{(1,j)}}{B_{(1,1)}} = \frac{\sigma_j(\mathcal{S}_\alpha^c)}{\sigma(\mathcal{S}_\alpha^c)} S_1 = r_j^S(\mathcal{S}_\alpha^c) S_1, \text{ where } \begin{cases} \sigma(\mathcal{S}_\alpha^c) = \sum_{\tau \in \Theta(S_1)} \pi(\tau) \neq 0 \\ \sigma_j(\mathcal{S}_\alpha^c) = \sum_{\tau \in \Theta(S_j)} \pi(\tau) \end{cases}$$

and $j = 2, \dots, N_\alpha$. It follows that $\sigma(\mathcal{S}_\alpha^c)$ is S-positive and $\sigma_j(\mathcal{S}_\alpha^c)$ is either a zero or S-positive element of $\mathbb{R}(\text{Con})[\mathcal{S}_\alpha^c]$. If the graph $G_{\mathcal{S}_\alpha}$ is strongly connected, then $\sigma_j(\mathcal{S}_\alpha^c) \neq 0$ for all j and any choice of S_j could be used instead of S_1 . Further:

Proposition 3.12. *A connected component $G_{\mathcal{S}_{\alpha,l}}$ of the graph $G_{\mathcal{S}_\alpha}$ is strongly connected if and only if $\sigma_j(\mathcal{S}_\alpha^c)$ is a non-zero rational function in $\mathbb{R}(\text{Con} \cup \mathcal{S}_\alpha^c)$ for all $S_j \in \mathcal{S}_{\alpha,l}$.*

The results shown above provide a proof of the following lemma.

Lemma 3.13. *If a substrate $S_t \in \mathcal{S}_\alpha^c$ is a variable in the rational function $r_j^S(\mathcal{S}_\alpha^c)$ for some $S_j \in \mathcal{S}_\alpha$, then there is $S_i \in \mathcal{S}_\alpha$ and $S_u \in \mathcal{S}_\alpha^c$, such that $S_i + S_t$ ultimately reacts to $S_j + S_u$ via \mathcal{Y}_α .*

After substitution of the value of S_j given in (3.11) into Y_k (3.2) we obtain

$$(3.14) \quad Y_k = r_k^Y(\mathcal{S}_\alpha^c) S_1,$$

where r_k^Y is either zero or an S-positive rational function in \mathcal{S}_α^c with coefficients in $\mathbb{R}(\text{Con})$. If $\widehat{G}_{\mathcal{Y}_\alpha}$ is strongly connected then this function is non-zero.

Conservation laws. The sum of the species concentrations in $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ is conserved. If the total amount $\bar{S}_1 = S_1 + \dots + S_{N_\alpha} + Y_1 + \dots + Y_{P_\alpha}$ is given, we obtain

$$\bar{S}_1 = (1 + r_2^S(\mathcal{S}_\alpha^c) + \dots + r_{N_\alpha}^S(\mathcal{S}_\alpha^c) + r_1^Y(\mathcal{S}_\alpha^c) + \dots + r_{P_\alpha}^Y(\mathcal{S}_\alpha^c)) S_1,$$

where the coefficient of S_1 is an S-positive element of $\mathbb{R}(\text{Con} \cup \mathcal{S}_\alpha^c)$ and thus,

$$S_1 = \bar{r}_1^S(\mathcal{S}_\alpha^c),$$

with \bar{r}_1^S an S-positive rational function in \mathcal{S}_α^c with coefficients in $\mathbb{R}(\text{Con} \cup \{\bar{S}_1\})$.

Further, if $\bar{S}_1 > 0$ then $S_1 \neq 0$ at steady state and $S_1 > 0$ for non-negative values of the substrates in \mathcal{S}_α^c . This remark and Proposition 3.12 imply:

Proposition 3.15. *A connected component $G_{\mathcal{S}_{\alpha,l}}$ of the graph $G_{\mathcal{S}_\alpha}$ is strongly connected if and only if any steady state solution satisfies $S_j \neq 0$ for all $S_j \in \mathcal{S}_{\alpha,l}$, and any total amounts $\bar{S}_l > 0$.*

By substitution of S_1 by \bar{r}_1^S , we obtain

$$(3.16) \quad Y_k = \bar{r}_k^Y(\mathcal{S}_\alpha^c) := r_k^Y(\mathcal{S}_\alpha^c) \bar{r}_1^S(\mathcal{S}_\alpha^c), \quad S_j = \bar{r}_j^S(\mathcal{S}_\alpha^c) := r_j^S(\mathcal{S}_\alpha^c) \bar{r}_1^S(\mathcal{S}_\alpha^c)$$

with \bar{r}_k^Y, \bar{r}_j^S either zero or S-positive rational functions in \mathcal{S}_α^c with coefficients in $\mathbb{R}(\text{Con} \cup \{\bar{S}_1\})$.

Proposition 3.17. *Assume that for each $l = 1, \dots, n_\alpha$, there exists a spanning tree of $G_{\mathcal{S}_{\alpha,l}, \mathcal{Y}_{\alpha,l}}$ rooted at some species S_{i_l} . Then, equations (3.5) are satisfied if and only if*

$$S_j = r_j^S(\mathcal{S}_\alpha^c) S_{i_l}, \quad S_j \in \mathcal{S}_{\alpha,l},$$

where r_j^S is zero or an S -positive rational function in \mathcal{S}_α^c with coefficients in $\mathbb{R}(\text{Con})$. Further, the conservation law $\bar{S}_l = \sum_{S_i \in \mathcal{S}_{\alpha,l}} S_i + \sum_{Y_k \in \mathcal{Y}_{\alpha,l}} Y_k$ is fulfilled if and only if

$$(3.18) \quad S_{i_l} = \bar{r}_{i_l}^S(\mathcal{S}_\alpha^c),$$

where $\bar{r}_{i_l}^S$ is an S -positive rational function in \mathcal{S}_α^c with coefficients in $\mathbb{R}(\text{Con} \cup \{\bar{S}_l\})$.

In Example (2.1), the graph G_{S_1} has two connected components: S_5 , which does not allow further eliminations, and $S_1 \xrightleftharpoons[b_{1,2}]{\tilde{b}_{2,1}} S_2$, which is strongly connected. Selecting S_1 as the non-eliminated species we obtain

$$S_2 = \frac{\tilde{b}_{2,1}}{\tilde{b}_{1,2}} S_1 = \frac{d_{1,2}}{b_{1,4}^2 \mu_{2,3}^2 S_3} S_1, \quad Y_1 = \frac{d_{1,2} \mu_{2,3}^1}{b_{1,4}^2 \mu_{2,3}^2} S_1 = \frac{d_{1,2} (b_{1,4}^2 + c_{2,1})}{b_{1,4}^2 c_{1,2}} S_1, \quad Y_2 = \frac{d_{1,2}}{b_{1,4}^2} S_1.$$

The total amount equations $\bar{S}_1 = S_5 + Y_3$ and $\bar{S}_2 = S_1 + S_2 + Y_1 + Y_2$ give:

$$\bar{S}_1 = S_5 (1 + \mu_{4,5}^3 S_4), \quad \bar{S}_2 = \frac{d_{1,2}}{b_{1,4}^2} \left(\frac{1}{\mu_{2,3}^2 S_3} + \frac{b_{1,4}^2 + c_{2,1}}{c_{1,2}} + 1 + \frac{b_{1,4}^2}{d_{1,2}} \right) S_1.$$

Let $\tilde{r}_1^S(S_3, S_4) = \bar{S}_2 \left(\frac{1}{\mu_{2,3}^2 S_3} + \frac{b_{1,4}^2 + c_{2,1}}{c_{1,2}} + 1 + \frac{b_{1,4}^2}{d_{1,2}} \right)^{-1}$; thus:

$$(3.19) \quad \begin{aligned} S_1 &= \frac{b_{1,4}^2}{d_{1,2}} \tilde{r}_1^S(S_3, S_4), & S_2 &= \frac{\tilde{r}_1^S(S_3, S_4)}{\mu_{2,3}^2 S_3}, & S_5 &= \frac{\bar{S}_1}{1 + \mu_{4,5}^3 S_4}, \\ Y_1 &= \frac{(b_{1,4}^2 + c_{2,1})}{c_{1,2}} \tilde{r}_1^S(S_3, S_4), & Y_2 &= \tilde{r}_1^S(S_3, S_4), & Y_3 &= \frac{\mu_{4,5}^3 \bar{S}_1 S_4}{1 + \mu_{4,5}^3 S_4}. \end{aligned}$$

Thus, all species are given as S -positive rational functions of S_3, S_4 in the coefficient field $\mathbb{R}(\text{Con} \cup \{\bar{S}_1, \bar{S}_2\})$.

3.4. Steady state equations. To summarize, at steady state the intermediate complexes \mathcal{Y} can be expressed as rational functions of the substrates \mathcal{S} and therefore eliminated. Further, provided a cut \mathcal{S}_α exists, the variables \mathcal{S}_α can be expressed as functions of $\mathcal{S}_\alpha^c = \mathcal{S} \setminus \mathcal{S}_\alpha$ and therefore also eliminated. For the latter statement, we make use of the conservation laws (with given total amounts) for the species in \mathcal{S}_α determined by the connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$.

Specifically, consider the steady state equations (3.5) for \mathcal{S}_α^c . Substituting the expressions in (3.14) and (3.11) for \mathcal{Y} and \mathcal{S}_α provides the steady states equations in terms of \mathcal{S}_α^c and the selected variables S_{i_l} (one for each connected component of $G_{\mathcal{S}_\alpha}$). Using (3.16), the steady states equations are given in terms of \mathcal{S}_α^c only. Let $\mathcal{S}_\alpha^c = \{S_{N_\alpha+1}, \dots, S_N\}$ and let $\Phi_u(\mathcal{S}_\alpha^c) = 0$ be the equation obtained from $\dot{S}_u = 0$ after elimination of \mathcal{Y} and \mathcal{S}_α and

removal of denominators. The denominators can be chosen to be S-positive and we can multiply the expressions by the denominators without changing the *positive* solutions.

Assume that the graph $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ has n_α connected components, and recall that each of them gives rise to only one conservation law (Proposition 3.10). Extend the set of conservation laws to a maximal set of $\dim(\Gamma^\perp)$ laws.

Theorem 3.20. *Consider a PTM system for which there exists a cut \mathcal{S}_α . Further, assume that each connected component of $G_{\mathcal{S}_\alpha}$ admits a rooted spanning tree. If total amounts \bar{S}_l are given for the n_α connected components of $G_{\mathcal{S}_\alpha, \mathcal{Y}_\alpha}$ and the $\dim(\Gamma^\perp) - n_\alpha$ additional conservation laws, then the non-negative steady states of the system with positive values for all substrates in \mathcal{S}_α^c are in one-to-one correspondence with the positive solutions to*

$$\Phi_u(\mathcal{S}_\alpha^c) = 0, \quad \bar{S}_l = \varphi_l(\mathcal{S}_\alpha^c)$$

for $u = N_\alpha + 1, \dots, N$ and $l = n_\alpha + 1, \dots, \dim(\Gamma^\perp)$.

Proof. We have shown that any non-negative steady state solution with positive values for all substrates in \mathcal{S}_α^c must satisfy these equations. For the reverse, consider a positive solution $s = (s_{N_\alpha+1}, \dots, s_N)$ to the equations $\Phi_u(\mathcal{S}_\alpha^c) = 0$ and $\bar{S}_l = \varphi_l(\mathcal{S}_\alpha^c)$. For $i = 1, \dots, N_\alpha$, define s_i through equation (3.18) and y_k , $k = 1, \dots, P$, through equation (3.2). For positive rate constants and positive total amounts, s_i, y_k are non-negative (because of the S-positivity of the rational functions defining them). By construction these definitions automatically ensure that the conservation laws with total amounts \bar{S}_l , $l = 1, \dots, N_\alpha$, are satisfied (see Proposition 3.17).

By Proposition 3.1, the values y_1, \dots, y_P satisfy (2.2) for all k and hence the steady state equations of the intermediate complexes are satisfied. By Proposition 3.17 the values s_1, \dots, s_{N_α} satisfy (3.5). Since the latter is just (2.3) after substitution of (3.2), we see that (2.3) holds as well. Since $\Phi_u(\mathcal{S}_\alpha^c) = 0$ is the steady state equation $\dot{S}_u = 0$ after substitution of (3.2) and (3.18), this equation is also satisfied and the same reasoning applies to the equation $\bar{S}_l = \varphi_l(\mathcal{S}_\alpha^c)$, $l > n_\alpha$. Thus, $S_i = s_i$ and $Y_k = y_k$ is a solution to the steady state equations and satisfy the conservation laws corresponding to the total amounts \bar{S}_l . \square

This theorem together with Proposition 3.2(iii) and Proposition 3.15 gives the following corollary.

Corollary 3.21. *Assume that \hat{G}_Y is strongly connected and that for all $S \in \mathcal{S}$ there exists a cut \mathcal{S}_α such that $S \in \mathcal{S}_\alpha$ and $G_{\mathcal{S}_\alpha}$ is strongly connected. Then, $S_i = 0$ or $Y_k = 0$ is not a steady state solution for any i, k . With the notation of Theorem 3.20, the non-negative steady states of the system are in one-to-one correspondence with the non-negative solutions to*

$$\Phi_u(\mathcal{S}_\alpha^c) = 0, \quad \bar{S}_l = \varphi_l(\mathcal{S}_\alpha^c)$$

for $u = N_\alpha + 1, \dots, N$ and $l = n_\alpha + 1, \dots, \dim(\Gamma^\perp)$.

In Example (2.1), $\dim(\Gamma^\perp) - n_\alpha = 1$ and only one conservation law is missing, $\bar{S}_3 = S_3 + S_4 + Y_1 + Y_2 + Y_3$. The elimination procedure leads to the steady state equations

consisting of $\dot{S}_3 = 0$ (Φ_3) and \bar{S}_3 (φ_3):

$$0 = \Phi_3(S_3, S_4) = -b_{1,4}^2 \tilde{r}_1^S(S_3, S_4) + \frac{b_{3,5}^3 \mu_{4,5}^3 \bar{S}_1 S_4}{1 + \mu_{4,5}^3 S_4}$$

$$\bar{S}_3 = \varphi_3(S_3, S_4) = S_3 + S_4 + \frac{b_{1,4}^2 + c_{2,1} + c_{1,2}}{c_{1,2}} \tilde{r}_1^S(S_3, S_4) + \frac{\mu_{4,5}^3 \bar{S}_1 S_4}{1 + \mu_{4,5}^3 S_4}.$$

Since the conditions of Corollary 3.21 are fulfilled, any non-negative solution of this reduced system provides a non-negative steady state of the PTM system. The steady states of the other species, $S_1, S_2, S_5, Y_1, Y_2, Y_3$, are found from (3.19). In this specific example, the first equation is easily transformed into a linear equation in S_3, S_4 , and hence either S_3 or S_4 can be eliminated as well, providing a polynomial equation in the remaining variables. In this case, S-positivity is not guaranteed.

In the example we intentionally selected \mathcal{S}_α to have the highest possible number of elements, since all these variables are subsequently eliminated. In Example (2.10), the cut $\mathcal{S}_\alpha = \{E, S_3, S_4\}$ allows us to eliminate three substrates and reduce the steady state equations to a system of three equations in three variables.

In some systems (see e.g. Section 4.2) there two different cuts $\mathcal{S}_\alpha, \mathcal{S}'_\alpha$ might exist, such that the union is not a cut, but still all variables in $\mathcal{S}_\alpha \cup \mathcal{S}'_\alpha$ can be eliminated. Thus, more species might be eliminated if different cuts are considered.

4. EXAMPLES

4.1. TG framework. In [17], the authors provide a linear elimination procedure for the special case in which the set of substrates is partitioned into two distinct sets. In their context, a *PTM system* (here called *TG system*) consists of three non-empty and disjoint sets of species called enzymes, substrates, and intermediate complexes:

$$\text{Enz} = \{E_1, \dots, E_L\}, \text{Sub} = \{S_1, \dots, S_N\}, \text{Int} = \{Y_1, \dots, Y_P\},$$

and a set of reactions $\text{Rct} = R_a \cup R_b \cup R_c$ with

$$R_a = \{E_i + S_j \xrightarrow{a_{i,j}^k} Y_k | (i, j, k) \in I_a\} \quad R_c = \{Y_i \xrightarrow{c_{i,j}} Y_j | (i, j) \in I_c\}$$

$$R_b = \{Y_k \xrightarrow{b_{i,j}^k} E_i + S_j | (i, j, k) \in I_b\}$$

for $I_a, I_b \subseteq \{1, \dots, L\} \times \{1, \dots, N\} \times \{1, \dots, P\}$ and $I_c \subseteq \{1, \dots, P\}^2$, such that (i) All chemical species are involved in at least one reaction; (ii) For every intermediate complex Y_k there is at most one enzyme $E_{\eta(k)}$, such that $(\eta(k), j, k) \in R_a \cup R_b$ for some j ; (iii) If two intermediate complexes Y_k, Y_v are 1-linked, then $E_{\eta(k)} = E_{\eta(v)}$. Further, the graph \widehat{G}_Y and each connected component of the graph G_{Sub} are required to be strongly connected. In particular, the assumption that \widehat{G}_Y is strongly connected implies that any Y_k ultimately reacts to $S_i + S_j$ for some i, j . This is our Assumption (ii) of a PTM system.

Essentially, they consider post-translational modification systems in which the enzymes are not allowed to be modified. Let $\mathcal{S} = \text{Sub} \cup \text{Enz}$, $\mathcal{S}_\alpha = \text{Sub}$, and $\mathcal{S}_\alpha^c = \text{Enz}$. Properties (i)-(iii) imply that \mathcal{S}_α is a cut. Note that $\mathcal{Y}_\alpha = \mathcal{Y}_\alpha^c = \mathcal{Y}$. Thus the framework developed here is an extension of the framework developed in [17].

By assumption (iii) the graph $G_{\mathcal{S}_\alpha, \mathcal{Y}}$ has L connected components that provide L conservation laws for the enzymes: $\bar{E}_i = E_i + \sum_{k|\eta(k)=i} Y_k$, for $i = 1, \dots, L$. With the notation of Lemma 2.13, $N_\alpha^c = n_\alpha^c = L$, $P_\alpha^c = 0$, so that $N_\alpha^c + P_\alpha^c - n_\alpha^c = 0$ and thus a set of independent conservation laws of a TG system can be derived from the non-interacting graphs of $G_{\mathcal{S}, \mathcal{Y}}$. Further, the form of R_a and R_b ensures that any non-interacting graph contains species either from Enz or Sub , but not both. Thus, all conservation laws are associated with a connected component either of $G_{\text{Enz}, \mathcal{Y}}$ or $G_{\text{Sub}, \mathcal{Y}}$.

It follows that if all intermediate complexes ultimately dissociate into an enzyme and a substrate, and each connected component of $G_{\mathcal{S}_\alpha}$ admits a rooted spanning tree, then elimination of the variables in $\mathcal{S}_\alpha \cup \mathcal{Y}$ reduces the steady state equations to L equations derived from the total amount of enzymes.

4.2. Signaling cascades. Our setting is well-suited to study elimination of variables in signaling pathways. Signaling pathways form a special type of PTM systems and an extension of TG systems to include some substrates that also act as enzymes.

Definition 4.1. A *signaling cascade* is a collection of TG systems R^1, \dots, R^n , with corresponding sets of species

$$\text{Enz}^i = \{E_1^i, \dots, E_{L_i}^i\}, \quad \text{Sub}^i = \{S_1^i, \dots, S_{N_i}^i\}, \quad \mathcal{Y}^i = \{Y_1^i, \dots, Y_{P_i}^i\}$$

and sets of reactions $\text{Rct}^i = R_a^i \cup R_b^i \cup R_c^i$, for $i = 1, \dots, n$, satisfying the following conditions:

- (i) $(\text{Enz}^i \cup \text{Sub}^i \cup \mathcal{Y}^i) \cap (\text{Enz}^j \cup \text{Sub}^j \cup \mathcal{Y}^j) = \{E_1^{i+1}\} = \{S_{N_i}^i\}$ if $j = i + 1$ and it is empty otherwise.
- (ii) For all i , each connected component of the graph G_{Sub^i} admits a spanning tree rooted at $S_{N_i}^i$.
- (iii) All intermediate complexes ultimately dissociate into two substrates.

Condition (i) implies that a signaling cascade consists of independent TG systems “joined” by only one substrate acting as an enzyme in the layer below. This description fits signaling pathways in which the signal is transmitted downstream. Condition (ii) ensures that the intermediate complexes can be eliminated.

Let $N = N_1 + \dots + N_n$, $L = L_1 + \dots + L_n$ and $\mathcal{S} = \bigcup_i \text{Enz}^i \cup \text{Sub}^i$. For each i , consider the subset $\text{Sub}^i \subset \mathcal{S}$. The associated set of intermediate complexes is $\mathcal{Y}_{\text{Sub}^i} = \mathcal{Y}^i \cup \{Y_k \in \mathcal{Y}^{i+1} \mid \eta(k) = S_{N_i}^i\}$, and Sub^i is closed (TG systems do not incorporate reactions $S_u \rightarrow S_j$ among substrates or enzymes). By definition, substrates in Sub^i do not interact and thus Sub^i is a cut.

For simplicity, we assume that the graph G_{Sub^i} is connected for each i . By Proposition 3.17, elimination of the variables in Sub^i provides the steady state relation

$$S_j^i = r_j^i(\text{Enz}^i) S_{N_i}^i, \quad S_j^i \in \text{Sub}^i \setminus \{S_{N_i}^i\}.$$

By Lemma 3.13, r_j^i depends on the species in Enz^i only: if $S_u^i + S_t$ ultimately reacts to $S_j^i + S_r$ for some species S_u^i in Sub^i and $S_r \in \mathcal{S} \setminus \text{Sub}^i$ via \mathcal{Y} , then since $S_j^i \neq S_{N_i}^i$, $S_t = S_r = E_\eta^i$ for some $E_\eta^i \in \text{Enz}^i$. Further, if $Y_k \in \mathcal{Y}_{\text{Sub}^i}$, we let $Y_k = r_k^Y(\text{Enz}^i) S_{N_i}^i$ be the corresponding rational function.

Conservation laws. Since G_{Sub^i} is connected and admits a rooted spanning tree, the sum of the species in the graph $G_{\text{Sub}^i, \mathcal{Y}_{\text{Sub}^i}}$ provides the only conservation law among the species in $\text{Sub}^i \cup \mathcal{Y}_{\text{Sub}^i}$. Thus, for each i , let a total amount \bar{S}_i be given. We have at steady state

$$(4.2) \quad \bar{S}_i = \sum_{S_j^i \in \text{Sub}^i} r_j^i(\text{Enz}^i) S_{N_i}^i + \sum_{Y_k \in \mathcal{Y}_{\text{Sub}^i}} r_k^Y(\text{Enz}^i) S_{N_i}^i.$$

For $i = n$, $S_{N_n}^n \notin \text{Enz}^n$, and so $S_{N_n}^n$ is expressed as a rational function in Enz^n .

Thus, if we let $\text{Enz} = \bigcup_i \text{Enz}^i$, we have that the species in $\mathcal{S} \setminus \text{Enz}$ are given as rational functions in Enz with coefficients in $\mathbb{R}(\text{Con})$. Condition (iii) implies that for $E \in \text{Enz}^i \setminus \{S_{N_i}^i\}$, $\{E\}$ is a cut with associated (connected) graph G_{E, \mathcal{Y}_E} . Thus, if the total amount \bar{E} is provided, the steady states must fulfill the equality

$$(4.3) \quad \bar{E} = E + \sum_{k|E=E_{\eta(k)}} Y_k = E + \sum_{k|E=E_{\eta(k)}} r_k^Y(\text{Enz}^i) S_{N_i}^i.$$

We conclude that the non-negative steady states of a signaling cascade are solutions to L equations in Enz with coefficients in $\mathbb{R}(\text{Con})$, provided that total amounts for Enz are given; that is, $\bar{S}_1, \dots, \bar{S}_{n-1}$ for the enzymes $S_{N_i}^i$, (4.2) and \bar{E}_η^i for $E_\eta^i \in \text{Enz} \setminus \{S_{N_1}^1, \dots, S_{N_{n-1}}^{n-1}\}$, (4.3).

Note that the number of conservation laws obtained in this way is $m = \sum_i L_i + 1$ (remember \bar{S}_n). Let $\epsilon = 1$ if n is even and 0 otherwise, and let $\epsilon^c = 1 - \epsilon$. The cuts provide all conservation laws: The graph associated to the cut

$$\mathcal{S}_\alpha = \bigcup_{i \text{ even}} \text{Sub}^i \cup \bigcup_{i \text{ odd}} \text{Enz}^i$$

has $n_\alpha = \epsilon + \sum_{i \text{ odd}} L_i$ connected components and thus, $n_\alpha^c = \epsilon^c + \sum_{i \text{ even}} L_i$. We have $N_\alpha = \sum_{i \text{ odd}} L_i + \sum_{i \text{ even}} (N_i - 1) + \epsilon$, and $N_\alpha^c = \sum_{i \text{ even}} L_i + \sum_{i \text{ odd}} (N_i - 1) + \epsilon^c$. Further, $\mathcal{Y}_\alpha = \mathcal{Y}$, so that $P_\alpha^c = 0$.

Let $\gamma = N_\alpha^c - n_\alpha^c = \sum_{i \text{ odd}} (N_i - 1)$. By Lemma 2.13, if there are γ independent terms in $\mathcal{S}_\alpha^c \cap \Gamma$, then all conservation laws come from non-interacting graphs. By hypothesis, for i even, the graph G_{Sub^i} has a spanning tree rooted at some node S_j . This means that for every $S_u \neq S_j$ in Sub^i , there exists a directed path $S_u \rightarrow S_{k_1} \rightarrow \dots \rightarrow S_{k_r} \rightarrow S_j$. By the conditions of a TG system and Lemma 3.9, an edge $S_{k_v} \rightarrow S_{k_s}$ implies that there is either a reaction $S_{k_v} \rightarrow S_{k_s}$, or $E + S_{k_v}$ ultimately reacts to $E + S_{k_s}$ via \mathcal{Y} . In either case, we see that $S_u - S_j \in \mathcal{S}_\alpha^c \cap \Gamma$ for all $S_u \neq S_j$ in Sub^i , implying that there are indeed γ independent vectors in $\mathcal{S}_\alpha^c \cap \Gamma$.

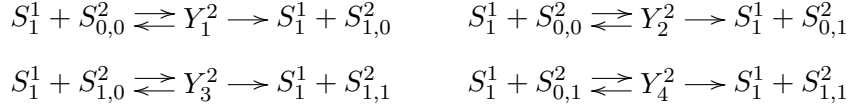
4.3. Biological examples.

MAPK signaling cascade. We consider the first two layers of the MAPK cascade: a two-layer cascade with one-site modification in the first layer and two-site modifications in the second layer. In the latter, dephosphorylation is considered sequential but this is not the case for phosphorylation [15].

The reactions of the system in the first layer are



accounting for phosphorylation and dephosphorylation, respectively, via a Michaelis-Menten mechanism. In the second layer we have the phosphorylation reactions

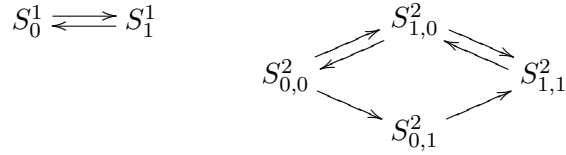


Dephosphorylation proceeds sequentially in the following way:



The sets of enzymes are $\text{Enz}^1 = \{E, F_1\}$, $\text{Enz}^2 = \{S_1^1, F_2\}$. The sets of substrates are $\text{Sub}^1 = \{S_0^1, S_1^1\}$, $\text{Sub}^2 = \{S_{0,0}^2, S_{1,0}^2, S_{0,1}^2, S_{1,1}^2\}$. The sets of intermediate complexes are $\text{Int}^1 = \{Y_1^1, Y_2^1\}$, $\text{Int}^2 = \{Y_1^2, Y_2^2, Y_3^2, Y_4^2, Y_5^2, Y_6^2\}$. We have $\text{Enz}^2 \cap \text{Sub}^1 = \{S_1^1\}$, so that the modified substrate in the first layer is a kinase of the next layer. The superindex denotes the layer, while the subindex denotes phosphorylation state (the presence of the phosphate group is represented by 1).

The components of the graph \hat{G}_Y are each of the intermediate complexes and are thus strongly connected. The graphs \hat{G}_{Sub^1} and \hat{G}_{Sub^2} are



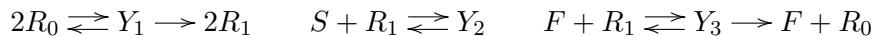
which are also strongly connected. The conservations laws (all derived from non-interacting graphs) are

$$\begin{aligned} \bar{E} &= E + Y_1^1 & \bar{S}_1 &= S_0^1 + S_1^1 + Y_1^1 + Y_2^1 + Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 \\ \bar{F}_1 &= F_1 + Y_2^1 & \bar{S}_2 &= S_{0,0}^2 + S_{1,0}^2 + S_{0,1}^2 + S_{1,1}^2 + Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 + Y_6^2 \\ \bar{F}_2 &= F_2 + Y_5^2 + Y_6^2 \end{aligned}$$

Therefore, if total amounts are provided, then the steady states of the two-layer cascade are found as solutions to a system of four polynomial equations in four variables, namely E, F_1, F_2, S_1^1 .

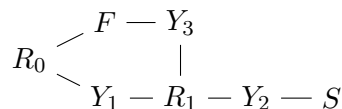
Receptor protein-tyrosine kinase. Receptor protein-tyrosine kinases (RPTK) are cell surface receptors linked to enzymes that phosphorylate their substrate proteins in tyrosine residues. The common mechanism for their activation is autophosphorylation following ligand-induced dimerization [2, §15]. The phosphorylated receptor serves as binding site to downstream signaling molecules, such as SH2 domain containing proteins. Further, the receptor can be dephosphorylated by several protein tyrosine phosphatases (PTP) [16].

A simple model describing the phosphorylation state of an RPTK is:



where R_0, R_1 stands for the unphosphorylated and phosphorylated RPTK respectively, S is a protein binding R_1 , and F is a PTP.

We have $\mathcal{S} = \{R_0, R_1, S, F\}$ and $\mathcal{Y} = \{Y_1, Y_2, Y_3\}$. Note that $\mathcal{S}_\circ = \{R_0, R_1\}$ are the self-interacting substrates and thus cannot be part of a cut. First of all, the intermediate complexes Y_k can be eliminated in terms of \mathcal{S} . The graph $G_{\mathcal{S}, \mathcal{Y}}$ is



The non-interacting graphs provide two conservation laws: $\bar{F} = F + Y_3$, and $\bar{S} = S + Y_2$, associated to the cut $\mathcal{S}_\alpha = \{F, S\}$. Thus, the substrates F, S can be eliminated, in fact from the conservation laws. We conclude that at steady state all species are described as rational functions of R_0, R_1 and the non-negative steady states are in one-to-one correspondence with the non-negative solutions to the equations corresponding to \dot{R}_0 and the remaining conservation law $\bar{R} = R_0 + R_1 + 2Y_1 + Y_2 + Y_3$.

ACKNOWLEDGMENTS

EF is supported by a postdoctoral grant from the “Ministerio de Educación” of Spain and the project MTM2009-14163-C02-01 from the “Ministerio de Ciencia e Innovación”. CW is supported by the Lundbeck Foundation, Denmark and the Leverhulme Trust, UK.

REFERENCES

- [1] D. Angeli, P. De Leenheer, and E. Sontag. Graph-theoretic characterizations of monotonicity of chemical networks in reaction coordinates. *J. Math. Biol.*, 61:581–616, Oct 2010.
- [2] G. M. Cooper and R. E. Hausman. *The cell*. ASM Press, Washington, fifth edition, 2009.
- [3] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: semiopen mass action systems. *SIAM Journal on Applied Mathematics*, 70(6):1859–1877, 2010.
- [4] Gheorghe Craciun and Martin Feinberg. Multiple equilibria in complex chemical reaction networks. I. The injectivity property. *SIAM J. Appl. Math.*, 65(5):1526–1546, 2005.
- [5] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.
- [6] M. Feinberg. Lectures on chemical reaction networks 1. *Notes*, pages 1–18, Jun 1980.
- [7] M. Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors i. the deficiency zero and deficiency one theorems. *Chem. Eng. Sci.*, 42(10):2229–68, 1987.
- [8] M. Feinberg and F. J. M. Horn. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Arch. Rational. Mech. Anal.*, 66(1):83–97, 1977.
- [9] E. Feliu, M. Knudsen, L. N. Andersen, and C. Wiuf. An algebraic approach to signaling cascades with n layers. *Bull. Math. Biol.*, 2011.
- [10] Jonathan L. Gross and Jay Yellen. *Graph theory and its applications*. Discrete Mathematics and its Applications (Boca Raton). Chapman & Hall/CRC, Boca Raton, FL, second edition, 2006.
- [11] R. Heinrich, B. G. Neel, and T. A. Rapoport. Mathematical models of protein kinase signal transduction. *Mol. Cell*, 9:957–970, May 2002.
- [12] C. Y. Huang and J. E. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U.S.A.*, 93:10078–10083, Sep 1996.
- [13] B. N. Kholodenko and M. R. Birtwistle. Four-dimensional dynamics of MAPK information processing systems. *Wiley Interdiscip Rev Syst Biol Med*, 1:28–44, 2009.

- [14] T. Krell, J. Lacal, A. Busch, H. Silva-Jimenez, M. E. Guazzaroni, and J. L. Ramos. Bacterial sensor kinases: diversity in the recognition of environmental signals. *Annu. Rev. Microbiol.*, 64:539–559, Oct 2010.
- [15] N. I. Markevich, J. B. Hoek, and B. N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.*, 164:353–359, Feb 2004.
- [16] A. Ostman and F. D. Bohmer. Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatases. *Trends Cell Biol.*, 11:258–266, Jun 2001.
- [17] M. Thomson and J. Gunawardena. The rational parameterization theorem for multisite post-translational modification systems. *J. Theor. Biol.*, 261:626–636, Dec 2009.
- [18] M. Thomson and J. Gunawardena. Unlimited multistability in multisite phosphorylation systems. *Nature*, 460:274–277, Jul 2009.
- [19] W. T. Tutte. The dissection of equilateral triangles into equilateral triangles. *Proc. Cambridge Philos. Soc.*, 44:463–482, 1948.
- [20] A. C. Ventura, J. A. Sepulchre, and S. D. Merajver. A hidden feedback in signaling cascades is revealed. *PLoS Comput. Biol.*, 4:e1000041, Mar 2008.